To appear in the 1998 *Symposium on Frontiers in Engineering*, sponsored by the National Academy of Engineering, Irvine, CA, September 1998.

Robotic Perception for Autonomous Navigation of Mars Rovers

Larry H. Matthies
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

The U.S. has been working on autonomous navigation of robotic vehicles since the 1960's, when the Jet Propulsion Laboratory (JPL) developed a prototype lunar rover for the Surveyor program. The earliest U.S. and Russian rovers were essentially teleoperated, which was acceptable for the few seconds of communication delay between Earth and the moon. In the 1970's, JPL began working on rovers for Mars, where communication latency of 10 to 20 minutes, due to round-trip light time, required a higher degree of autonomy on the rover if exploration was to be efficient [1]. Key technical barriers were relatively poor 3-D sensing of the environment, lack of accurate means to keep track of the position of the rover as it travelled, and the low performance of onboard computers. This paper surveys key progress in sensors, algorithms, and processors that has alleviated these specific barriers, thereby enabling practical autonomous navigation.

Background

Research on robotic vehicles for terrestrial applications accelerated in the late 1970's and throughout the 1980's. The Autonomous Land Vehicle (ALV) Program, funded by DARPA from 1984 to 1988, developed a vehicle the size of a large bread truck that used a scanning laser range-finder to locate and avoid obstacles while driving cross-country at speeds on the order of 2 mph [2]. Progress in this area was retarded by the size, cost, and power consumption of the laser range-finder and by the size and cost of the onboard computers. Meanwhile, in the more structured environment of autonomous lane-following on highways, researchers in Germany were able to achieve driving speeds of up to 60 miles an hour for a Mercedes van that used onboard cameras and computers to track lane markings [3]. This application was helped by the strong geometric constraints available from knowledge of highway design and by the relatively low amount of computation required to track lane markings in imagery. Nevertheless, this effort made a significant contribution to the field by demonstrating effective application of system dynamics models and Kalman filtering in a computer vision system for autonomous navigation.

Real-time 3-D Perception for Mars Rovers

A key breakthrough in 3-D perception for autonomous, cross-country navigation came in 1990, when JPL developed efficient, reliable algorithms for estimating elevation maps in real time from stereo image pairs using compact, commercial processors onboard a prototype Mars rover [4]. 3-D sensing with stereo cameras has advantages over laser scanners for Mars rover applications because it is easier to make the sensor mechanically robust enough to survive the rigors of launch and landing. This "stereo vision" approach computes range by triangulation, given matching features in the two images. The prevailing approach to stereo matching in the computer vision community at that time, popularized by David Marr at MIT [5], was to first extract edges from both images, then to match the edges found in the left and right image. The edge detection process reduces the amount of information to be processed in the matching stage, which is what made this approach appear attractive; however, by definition it produces "sparse" range data, because it measures range only where prominent edges are found in the imagery. For Mars rover applications, this would mean that range data would only be found around the outlines of high-contrast rocks. However, to do effective traverse planning, it was preferable to sample the elevation of the terrain more densely, so as to be guaranteed to find small obstacles that might trap a wheel.

The heart of JPL's innovation was to develop an efficient stereo matching algorithm based on area correlation, which was able to produce reliable range measurements at almost every pixel in the image for applications like Mars rovers. Most components of this algorithm had been developed previously for different problems; hence, the key contribution was to recognize how to put them together to produce a fast, reliable solution to this problem. The first step in this algorithm is a process called rectification, which resamples both images in such a way that corresponding image features lie on corresponding scanlines in the two images; this reduces the search for matching features to a 1-D search along the corresponding scanlines. The second step applies a bandpass filter to each image to compensate for overall differences in brightness between images from the two cameras. The third step finds corresponding features by evaluating a least squares similarity criterion for a small image patch from the left image at several trial match positions along the scanline in the right image. This evaluation is performed for each pixel in the left image to produce a range estimate at each pixel; the large number of correlations to perform makes this the bottleneck step in the whole algorithm. In straightforward implementations, the number of arithmetic operations needed to evaluate the least squares criterion is 3N, not including addressing operations, where N is the number of pixels in the image patch used for matching; N is typically around 50 (ie. a 7×7 patch). However, there exists an incremental technique that maintains intermediate results to reduce the cost to 6 operations per trial match position, independent of the size of the image patch used for matching. This was key to making the entire process practical for real-time implementation onboard a robotic vehicle.

The first incarnation of a stereo vision system using this algorithm required nearly 10 slots in a 6U VME card cage (ie. on the order of 1 cubic foot) and 100 to 200 Watts of power. This sytem produced about 1000 range measurements per second; this enabled the JPL Mars rover prototype "Robby" to drive autonomously over a 100-meter cross-country course in 4 hours in September of 1990 [4]. This was a significant "first" for a robotic vehicle.

Further Development of Real-time 3-D Perception

This technology was picked up by DARPA and the U.S. Army for use in research programs aimed at developing unmanned ground vehicles for military reconnaissance applications [6]. By 1996, the speed of stereo vision systems had increased to about 30,000 range measurements per second, with faster computers that occupied slightly less space. This enabled semi-autonomous HMMWV's to execute rudimentary reconnaissance missions covering about 2 miles of open terrain at speeds up to 5 mph [7].

Since 1996, the advent of general purpose microprocssors with limited vector processing capability has enabled substantial speed improvements and size reductions for stereo vision systems. The MMX feature in the Intel Pentium is the best-known example of such a capability; it allows up to four 16-bit integer operations to be performed in parallel. Stereo algorithms implemented on these processors can now perform about 700,000 range measurements/second [8]. Lower performance systems based on the same algorithm but different CPU's have been built on circuit cards that fit in the palm of the hand – including the CPU and both cameras. This highlights another recent development that is enabling compact, low cost computer vision systems for robotic vehicles: the introduction of low-power CMOS imagers with clocking, control, and analog-to-digital conversion functions fully integrated on-chip. Forthcoming advances in vector processing for general purpose CPU's will accelerate these trends; for example, Motorola has announced a vector-processing extension to the PowerPC architecture that will allow 16 byte-wide arithmetic operations to proceed in parallel. Within a year, this will enable stereo vision systems to produce on the order of 2 million range measurements per second with a single microprocessor; that is, 256x256 range measurements per frame at full video rate (30 frames/second). Compared to 1990, this is a speed increase of three orders of magnitude with a simultaneous reduction in size and power dissipation of one order of magnitude.

These advances are enabling a suite of new applications of robotic vehicles. In 1999, a robotic vehicle carrying stereo cameras is to enter Chornobyl to attempt to create a 3-D model of the interior to facilitate further clean-up efforts. JPL's stereo algorithm will enable future Mars rovers to explore several kilometers, in comparison with the 100 meters or so covered by the Sojourner rover in the summer of 1997. DARPA and the U.S. Army are also continuing to use this technology for further development of military robotic vehicles. For example, the "Demo III" program managed by the Army Research Laboratory aims to enable autonomous cross-country navigation at 20 mph by 2001, for a robotic vehicle the size of a large desk. The DARPA Tactical Mobile Robot Program is currently funding development of robotic vehicles the size of a large briefcase for reconnaissance applications in urban warfare. Both of these programs will employ stereo vision among their sensor suites for autonomous navigation and will depend on the aforementioned advances in algorithms, low-power CMOS imagers, and high performance embedded CPU's to provide the increased speed and smaller size required.

Limitations and Approaches to Solutions

These advances have produced a viable solution to real-time 3-D perception for robotic vehicles operating during the day in barren or semi-arid terrain. Limitations that arise as we push for broader applicability include:

- For military applications, operability at night is essential. It appears that stereo vision with thermal infrared imagery works quite well, though thermal cameras are currently very expensive. Two-axis scanning laser range-finders work well at night, but are also still large and expensive.

- Stereo vision fails in textureless environments, such painted walls in indoor mobile robot applications. This can be solved by adding low cost, low resolution active sensors (eg. sonar) or compact, single-axis scanning laser range-finders to sense the floorplan of a room.

- For terrestrial applications, robotic vehicles need to perceive both the 3-D geometry and the composition of terrain (eg. to discriminate traversable vegetation from non-traversable rocks). For some basic discriminations, viable solutions are in hand; in particular, live vegetation is easily distinguished from soil and rocks using visible and near-infrared imagery [7]. Other discriminations are still poorly solved, especially for real-time applications, such as distinguishing dead vegetation from soil. For some of these cases, we are studying the use of image texture for terrain classification.

- In addition to obstacle detection, position estimation is a major part of the autonomous navigation problem. Although GPS can largely solve this problem outdoors on Earth, it is still an important problem for indoor robot applications and in planetary exploration. Visual feature tracking with stereo cameras has been employed successfully for robot motion estimation and for terminal guidance to human-designated objectives [9, 10]. A number of methods are under development that use maps together with images and other sensors for various forms of landmark recognition [11, 12, 13]. Some of these methods are fairly mature for Earth applications now; methods suitable for Mars rovers will likely come to maturity and be deployed over the next 5 to 7 years.

- In terrestrial applications, autonomous navigation among other moving objects requires a significant extension of perception, planning, and local world modeling capabilities beyond that addressed above. Much work is in progress on this problem, using sonar, scanning laser range-finders, and imagery; however, we will stop short of surveying it here.

Predictions

Robotic perception systems for Mars rovers should enable autonomous navigation of up to a kilometer or more from the lander by the year 2003. Laser range-finders and image-based feature tracking and landmark recognition algorithms are expected to be used for autonomous precision landing on a comet in less than 10 years from now. Within 10 years, it is also possible that unmanned ground vehicles will be sufficiently mature to proceed with fielding them for selected military applications. The cost of such systems may also be low enough, and the capability high enough, to support commercialization for some civil applications. Potential commercial applications of such technology include autonomous material transportation and collision avoidance sensors for smart passenger vehicles. Applications for the computer vision technology also exist outside of autonomous navigation, such as in PC-based camera systems that use 3-D shape and motion sensing capabilities to enhance video conferencing and human-computer interfaces. In fact, commercial imaging applications are part of what is driving the rapid progress in low-power

CMOS imagers, low-power embedded processors, and vector processing extensions to general purpose microprocessor architectures. These advances will lead to computer vision systems within the next 10 years that will make robotic vehicles and the vision systems themselves cost-effective for new applications. Finding and developing these applications will be an exciting opportunity for engineers in the next decade.

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, with support from the National Aeronautics and Space Administration, Defense Advanced Research Projects Agency, and the Joint Robotics Program of the Office of the Secretary of Defense.

References

[1] D. A. O'Handley. Scene analysis in support of a Mars rover. *Computer Graphics and Image Processing*, 2:281–297, 1973.

[2] K. E. Olin and D. Y. Tseng. Autonomous cross-country navigation. *IEEE Expert*, 6(4):16–32, August 1991.

[3] E. D. Dickmanns and B. Mysliwetz. Recursive 3-D road and relative ego-state recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):199–213, 1992.

[4] L. H. Matthies. Stereo vision for planetary rovers: stochastic modeling to near real-time implementation. *International Journal of Computer Vision*, 8(1):71–91, July 1992.

[5] D. Marr. *Vision*. W. H. Freeman and Company, San Francisco, 1982.

[6] E. G. Mettala. The OSD Tactical Unmanned Ground Vehicle Program. In *Proc. DARPA Image Understanding Workshop*, pages 159–172. Morgan Kaufmanm Publishers, January 1992.

[7] L. Matthies, A. Kelly, T. Litwin, and G. Tharp. Obstacle detection for unmanned ground vehicles: a progress report. In G. Giralt and G. Hirzinger, editors, *Robotics Research: Proceedings of the 7th International Symposium*. Springer-Verlag, 1996.

[8] K. Konolige. Small vision systems: hardware and implementation. In *Eighth International Symposium on Robotics Research*, October 1998.

[9] L. H. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):239–248, June 1987.

[10] D. wettergreen, H. Thomas, and M. Bualat. Initial results from vision-based control of the Ames Marsokhod rover. In *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, September 1997.

[11] F. Cozman and E. Krotkov. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. In *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision*, December 1996.

[12] F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2D range scans. *Journal of Intelligent and Robotic Systems*, 18:249–275, 1997.

[13] L. Matthies, C. Olson, G. Tharp, and S.. Laubach. Visual localization methods for Mars rovers using lander, rover, and descent imagery. In *Proc. International Symposium on Artificial Intelligence, Robotics, and Automation in Space (i-SAIRAS)*, July 1997.